

# Instabilities and Trust: Basic Research Data Under Construction

Sharon Traweek

Gender Studies & History Departments

University of California, Los Angeles

Data 2012, January 26, 2012

# Funding, 2009-2014

- \* The Transformation of Knowledge, Culture, and Practice in Data-Driven Science: A Knowledge Infrastructures Perspective, Award# 20113194, Alfred P. Sloan Foundation, 2012-2014
- \* EAGER: Knowledge and Data Transfer: the Formation of a New Workforce, Award# 1145888, National Science Foundation, September 2011-August 2012
- \* The Data Conservancy: A Digital Research and Curation Virtual Organization, Award# 0830976, National Science Foundation, August 2009-July 2012

# Collaborators

- Christine Borgman, PI, UC Presidential Chair in Information Studies, UCLA

<http://is.gseis.ucla.edu/cborgman/>

<http://works.bepress.com/borgman/>

- Sharon Traweek, co-PI, Gender Studies & History, UCLA

<http://www.history.ucla.edu/traweek/>

[http://www.womensstudies.ucla.edu/faculty\\_traweek.html](http://www.womensstudies.ucla.edu/faculty_traweek.html)

- Laura Wynholds and Ashley Sands, Graduate Student Researchers, Information Studies, UCLA

<http://is.gseis.ucla.edu/academics/degrees/phd/students.htm>

- Former member: David Fearon, Entrepreneurial Library Program, Johns Hopkins University

<http://www.library.jhu.edu/elp/Team/fearon.html>

# Knowledge Infrastructures in Astronomy Data, 2009-2014

<http://knowledgeinfrastructures.gseis.ucla.edu/index.html>

Research Questions:

What are the current practices in data design, collection, access, use, reuse, revision, sharing, and curation for basic research in astronomy?

What are the variations in those practices?

Who is doing this work?

How is all this changing and why ?

# Our Research Methods

- Interviews and oral histories
- Multi-sited ethnographies at universities, laboratories, and conferences
- Following databases: formation, sharing access, revisions, curation
- Following data from design and collection to publication and citation.

Our multi-sited ethnographic & oral history research includes large-scale database-driven astronomy projects (such as SDSS & LSST)

1. Large international, distributed teams (>100)
2. Huge archived digital databases from sky surveys
3. Data used to generate images for research
4. Increased access to research data
5. Increased workforce gathering & interpreting data
6. Changing organizational and funding infrastructure
7. Support by government/industry/university sectors
8. Two primary examples: SDSS & LSST

# Basic Research

- State-of-the-art project design, equipment, data collection, databases, and analytic tools are not standardized and always under repair.
- Stable equipment, data, and analytic tools are used to calibrate backgrounds against which new kinds of data can be recognized and the long process of data evaluation can proceed.
- Basic research communities know how to build knowledge in the context of this mixture of stability and instability in design, equipment, data, and analysis. They develop different approaches to answering shared questions.

# Sloan Digital Sky Survey (SDSS)

<http://www.sdss.org/>

<http://www.sdss.org/collaboration/>

- Sloan I 2000–05; Sloan II 2005–08; Sloan III 2008-14
- Funded by Sloan Foundation
- Telescope at Apache Point, NM, USA
- Partnership with Microsoft: Skyserver Database

<http://skyserver.sdss.org/dr1/en/skyserver/>

- \* SDSS database access rates: 952,336 monthly 2009-11 <http://skyserver.sdss.org/log/en/traffic/>



# Large Synoptic Survey Telescope

<http://www.lsst.org>

- LSST is the next big project in astronomy
- Scheduled to be ready by 2014
- Telescope at Cerro Pachón, Chile
- Collaborations now forming
- Includes many physicists
- Partnership with Google to download raw data on internet every three nights

# One astronomer's comment on access to LSST data

“LSST is the first large facility ... where the data [will be] public in thirty seconds. Every high school student in the world... [will have] the same access to the data that I do.... That means, if you're interested in something, you can get the data.”

# Variation in Astronomy Research Practices

Types of Observatories:

Ground & Space

Data Deposit Sites:

Facilities & Desktops

Investigators based at:

Universities,

Observatories, &

Space Research Facilities

Funding sources:

Govt agencies

Private foundations

Topic & Forms of Inquiry:

Extra-terrestrial signals

Star formation

NASA archives

Exoplanets

Galaxy evolution

X-ray space telescopes

Gamma-ray bursts

**Radio waves**

**Microwave &  
Infrared**

**Ultraviolet**

**X rays**

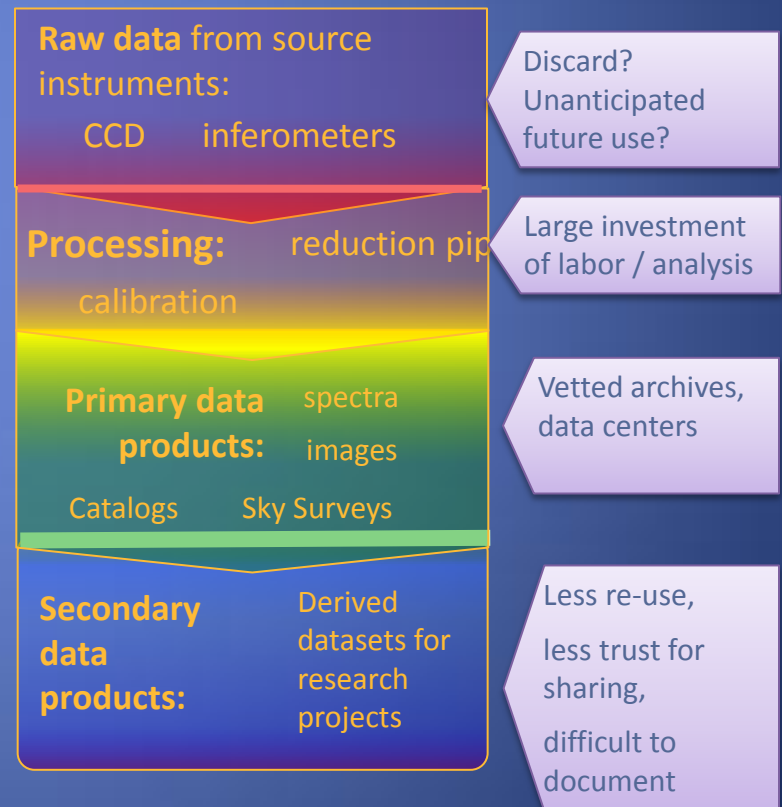
**Gamma rays**

Sky Surveys

# Findings: Building Trust in Data

- Building trust in **primary data**
  - Iterative calibration & testing
  - Vetted data systems and services
  - Human expertise
- Building trust in **secondary data**
  - Difficult to discover
  - Inconsistent documentation
  - N.I.H.

## ► Astronomy



# Findings: stable & unstable data coexist in basic research

Data curation for basic research must accommodate the on-going daily work in basic research with both stable and instable data (data undergoing the long process of evaluation), including those in large-scale data sets.

# National Science Foundation (NSF)EAGER Program

- EArly-concept Grants for Exploratory Research  
...support exploratory work in its early stages on untested, but potentially transformative, research ideas or approaches. NSF is in the process of implementing a new emphasis on transformative research ...

[http://www.nsf.gov/pubs/policydocs/pappguide/nsf09\\_29/gp\\_g\\_2.jsp#IID2](http://www.nsf.gov/pubs/policydocs/pappguide/nsf09_29/gp_g_2.jsp#IID2)

# Women and Minority Astronomers Strategic Engagement with Distributed, Multi-Disciplinary Collaborations and Large Scale Databases

- An NSF OCI funded research project 2009-2012
- The goal is to understand how gender, ethnicity, class, and nationality in the astronomy workforce intersects with changing knowledge-making practices.

# Collaborators

- PI: Sharon Traweek, UCLA, anthropologist & historian of 20c physical sciences (Europe, Japan, & US)
- Co-PI: Jarita Holbrook, University of Arizona, PhD, Astronomy & Astrophysics. Ethnographer of cultural astronomy practices (Fiji, Tunisia, UK, & US)
- Postdoctoral Research Associate: Reynal Guillen, UCLA Chicano Studies Research Center; BS/MS in astronomy & space physics & PhD, history of science. Ethnographic historian of minorities in science & engineering in southwestern US



## 3 Graduate Student Research Assistants

- Diane Yu Gu (from China), Doctoral Candidate, Comparative Higher Education, UCLA: mentoring women grad students in physical sciences
- Luis Felipe R. Murillo (from Brazil), Graduate Student, Cultural Anthropology, UCLA: networks among open source groups in Brazil, Japan, & US
- Brad Fidler (from Canada), UCLA PhD, big pharma marketing strategies with big data